

Medical RAG Chatbot - Explainable & Citation-Based Q&A System

Overview

This project is a Retrieval-Augmented Generation (RAG) based Medical Chatbot that provides accurate, explainable, and citation-supported answers to healthcare-related questions. It retrieves information from trusted medical sources and uses a language model to generate concise, evidence-based responses with clear references.

This chatbot is for informational and educational purposes only. It is not a substitute for professional medical advice

What is RAG?

Retrieval-Augmented Generation (RAG) combines:

1. Retrieval - fetching relevant documents or text chunks from verified medical sources (like WHO, PubMed, CDC).
2. Generation - using an LLM (e.g., GPT or LLaMA) to compose a human-readable answer based only on retrieved evidence, with citations for transparency.

This approach enhances:

- Accuracy
- Explainability
- Traceability (Citations & Sources)

System Architecture

graph TD;

A[User Query] --> B[Retriever: Vector DB (FAISS)];

B --> C[Relevant Text Chunks];

C --> D[Generator: LLM];

D --> E[Answer + Explanation + Citations];

E --> F[Frontend UI / Chat Interface];